Towards a system for the semi-automatic annotation of eye gaze data in face-to-face interactions

Stijn De Beugher, Geert Brône, and Toon Goedemé

Studies on language and cognition in interaction increasingly focus on the role of eye gaze as an important signal in interaction management, reference and grounding (Rossano 2012, Jokinen 2010, Bailly et al. 2010, Richardson et al. 2009). Interlocutors may use eye gaze as a means to take, hold or give the floor in conversation (turn management), to refer to objects or persons in the conversational space (gaze cueing) or to give and elicit feedback (grounding). The use of non-intrusive eye-tracking technology (like eye-tracking glasses or table-top systems) has proven to be an invaluable resource for obtaining detailed information on the distribution of visual attention of multiple participants simultaneously (Jokinen 2010, Oertel & Salvi 2013, Brône & Oben 2015, Oben 2015, Holler & Kendrick 2015). This generates a picture of the role of eye gaze in interactional dynamics, with speakers and addressees displaying different gaze patterns.

One of the key challenges in the use of mobile eye-tracking technology, however, resides in the processing and annotation of the obtained data stream. To date, the annotation process is largely manual, which is time-consuming and labour-intensive. Part of this work, however, can be automatized using recognition algorithms from vision technology (De Beugher, Brône & Goedemé 2013). In this paper, we present one such system for the semi-automatic recognition of human faces, torso and hands, thus providing a first categorization of targets on which the gaze data of the eye-tracking systems can be mapped. In other words, the system analyses data captured by the scene camera of the eye-tracker, calculates scores for the annotation classes (face, torso, hands), and then searches for matches between gaze coordinates and annotations (i.e. are there gaze fixations on faces, torsos and hands?).

The approach we present in this talk partly builds on previous work on the detection of human torso and faces (De Beugher, Brône & Goedemé 2013) but has been improved for the computational cost of the algorithms. The detection of hands, on the other hand, is based on an accurate segmentation in combination with advanced tracking mechanisms and a validation of human poses. The algorithms we used are embedded in a semi-automatic tool, which calculates the confidence of the hand detections. If the confidence drops below a certain threshold, the automatic analysis is halted and the user is asked for manual annotation. After this intervention, the system automatically continues the processing of the remaining frames. Using such an approach results in a highly accurate system at a minimal cost of manual interventions (average accuracy: 90%, manual interventions: 1,7% of the annotations, average processing time per frame: 150ms). As a last step, we map our detections on the gaze data and export those data to an ELAN compatible file, making the tool integratable with existing annotations.

References

Bailly, G., Raidt, S. & Elisei, F. (2010) Gaze, conversational agents and face-to-face communication. *Speech Communication - special issue on Speech and Face-to-Face Communication* 52, 598–612.

Brône, G. & Oben, B. (2015) InSight Interaction. A multimodal and multifocal dialogue corpus. *Language Resources and Evaluation* 49-1, 195-214.

De Beugher S., Brône G. & Goedemé T. (2013). Object recognition and person detection for mobile eye-tracking research. A case study with real-life customer journeys. In *Proceedings of the First International Workshop on Solutions for Automatic Gaze Data Analysis 2013* (SAGA 2013).

De Beugher S., Brône G. & Goedemé T. (2015). Semi-automatic hand detection - a case study on real life mobile eye-tracker data. In: *Proceedings of VISAPP*, pages 121-129.

Holler, J. & Kendrick K. H. (2015). Unaddressed participants gaze in multi-person interaction: optimizing recipiency. *Frontiers in Psychology* 6-98.

Jokinen, K. (2010). Non-verbal signals for turn-taking & feedback. In: *Proceedings of 7th International Conference on Language Resources & Evaluation*.

Oertel C. & Salvi G. (2013). A Gaze-based Method for Relating Group Involvement to Individual Engagement in Multimodal Multiparty Dialogue. In: *Proceedings of the 15th ACM International Conference on Multimodal Interaction ICMI.* Sydney, Australia.

Richardson, D., et al. (2009). Conversation, gaze coordination & beliefs about context. *Cognitive Science* 1468-1482.

Rossano, F. (2012). Gaze in Conversation. In: J. Sidnell & T. Stivers (eds.), *The Handbook of Conversation Analysis*. London: Wiley, 308-329.